

# Classification of Red Blood Cell Images Using a Neural Network

Samavedam A. Krishna<sup>1</sup>, MD; Joseph A. Orr<sup>2</sup>, PhD; Dwayne R. Westenskow<sup>2</sup>, PhD

<sup>1</sup>Department of Medical Informatics, <sup>2</sup>Department of Anesthesiology - Bioengineering Division  
University of Utah, Salt Lake City, Utah

A back-propagation artificial neural network (ANN) has been used to classify images of red blood cells (RBC) based on morphological features. The sensitivity and specificity of the ANN for correctly classifying four categories of RBC have been obtained, at different levels of image noise and with increasing training of the ANN.

## MATERIALS AND METHODS

Blood smears from four persons were photographed at 400x magnification; one person had normal cells (NC), one had microcytic hypochromic (MH) anemia, one had sickle cell (SC) anemia and one had hereditary spherocytosis (HS). The photomicrographs were transferred to computer media as 8-bit greyscale image files using a slide digitizer. The images were further manipulated using an image processing program, (NIH Image). Corrections were made for variation in background intensity of the images. Individual RBC were manually cropped from the photomicrograph images into separate files of 75 x 75 pixel size. From each RBC image, 5 values - the area, mean pixel density, standard deviation of pixel densities, integrated density and modal density - were obtained for pixels with greyscale values above the image noise threshold; two thresholds were used. These parameters were chosen for ease of acquisition and because they appeared to be independent descriptors of RBC morphology. Each set of 5 values constituted a pattern for presentation to the artificial neural network. At each threshold, 61 RBC were measured, with 14 NC, 16 MH, 13 SC and 18 HS cells, resulting in 61 patterns. For training the ANN, patterns at the same noise threshold were presented, as far as possible, in the order of one pattern per category following the other. As an important data pre-processing step prior to training and testing the ANN, all 61 values for each parameter were normalized to lie between -1 and +1.

Dartnet, a public-domain artificial NN simulation program, was used to classify the patterns. The ANN had 3 layers of neurons, with 5 neurons each in the input and hidden layers, and 4 in the output layer. Each input layer neuron received one of the 5 feature values in a pattern. Neuron outputs were passed on to all neurons of the next layer. For hidden and output layer neurons, outputs were calculated as logistic functions of summated inputs. Each of the 4 output layer neurons corresponded to one RBC diagnosis - determined by a high output value for one of the

neurons and low values for the other three. The ANN was trained by modifying neuron interconnection weights by back-propagation of error between actual ANN outputs and desired outputs for any training pattern. The ANN was trained and tested by cross-validation. One pattern of each category would be withheld during training and the ANN would be trained on the remaining 57 patterns of normalized values. After every 50 cycles of the training pattern set, the 4 test patterns would be presented to the ANN. Outputs in response to each of the 4 test patterns would be noted and the process continued till 2000 training cycles.

## RESULTS

After all 61 patterns had been tested, the results were reorganized as true positive, false positive and false negative with respect to the true diagnoses for the test patterns. These results were then further compiled to generate the true positive rate (TPR) and false positive rate (FPR) for diagnosing each category of cells at each of the two thresholds. Using these values, the diagnostic performance of the ANN has been expressed in terms of sensitivity (= TPR) and specificity (= 1 - FPR) for diagnosis of each category of cells at each threshold.

The sensitivity and specificity for diagnosing the four categories of cells, at both thresholds (noise levels), stabilized after about 500 to 600 training cycles. In some cases, there was a marginal drop in sensitivity and/or specificity with increasing number of training cycles. At a threshold of 45 (greater noise), the best sensitivities were 0.64, 0.69, 0.77 and 0.94; the corresponding specificities were 0.96, 0.65, 0.98 and 1.0 respectively for NC, MH, SC and HS cells. At a threshold of 128 (lesser noise), the best sensitivities were 0.79, 0.69, 0.92 and 1.0; the corresponding specificities were 0.98, 0.93, 0.98 and 1.0 respectively for NC, MH, SC and HS cells.

## CONCLUSION

Adequate to high sensitivity and specificity for identification of RBC images by an ANN using only a few image descriptors (5 parameters) can be achieved once regions of interest have been defined. The need for only a few training cycles is probably the result of pre-processing of data (by normalization); this contrasts with similar previous studies in which very large numbers of training iterations were used.